CY=JP DATE=19970610 KIND=A
PN=09149790

NOVEL SERINE PROTEASE
[Shinki serin puroteaaze]

Nobuo Tsuruoka et al.

UNITED STATES PATENT AND TRADEMARK OFFICE
Washington, D.C. September 2000

Translated by: Diplomatic Language Services, Inc.

PUBLICATION COUNTRY	(19): JP
DOCUMENT NUMBER	(11): 09149790
DOCUMENT KIND	(12): A (13):
PUBLICATION DATE	(43): 19970610
PUBLICATION DATE	(45):
APPLICATION NUMBER	(21): 08212196
APPLICATION DATE	(22): 19960724
ADDITION TO	(61):
INTERNATIONAL CLASSIFICATION	(51): C12N 15/09; C07H 21/04; C07K 14/47; C12N 1/21; 5/10; C12N 9/52; //A61K 38/46; (C12N 15/09; C12R 1:91); (C12N 1/21; C12R 1:19); (C12N 9/52; C12R 1:19);(C12N 9/52; C12R 1:91)
DOMESTIC CLASSIFICATION	(52):
PRIORITY COUNTRY	(33): JP
PRIORITY NUMBER	(31): 07275105
PRIORITY DATE	(32): 19950929
INVENTOR	(72): TSURUOKA, NOBUO; YAMASHIRO, KYOKO; TSUJIMOTO, MASAFUMI; YAMAGUCHI, NOZOMI
APPLICANT	(71): SUNTORY LTD.
TITLE	(54): NOVEL SERINE PROTEASE
FOREIGN TITLE	[54A]: SHINKI SERIN PUROTEAAZE

Title of the Invention /1

Novel serine protease

Claims /2

(Claim 1) Novel serine protease or its partial peptide comprised by including the amino acid sequence from amino acid 1 to 223 shown by sequence number: 3 (where the leucine of amino acid 1 may be missing), the amino acid sequence from amino acid 1 to 233 shown by sequence number: 4, or the amino acid sequence from amino acid 1 to 241 shown by sequence number: 5.

(Claim 2) DNA that codes a novel serine protease or its partial peptide comprised by including the amino acid sequence from amino acid 1 to 223 shown by sequence number: 3 (where the leucine of amino acid 1 may be missing), the amino acid sequence from amino acid 1 to 233 shown by sequence number: 4, or the amino acid sequence from amino acid 1 to 241 shown by sequence number: 5.

(Claim 3) DNA described in Claim 2 comprised by including the nucleotide sequences of nucleotides 219 to 887 and nucleotides 222 to 887 of sequence number: 3, the nucleotide sequence of nucleotides 1 to 699 of sequence number: 4, or the nucleotide sequence of nucleotides 1 to 723 of sequence number: 5 or its partial peptide.

(Claim 4) Recombinant vector comprised by including the DNA described in Claim 2 or 3.

(Claim 5) Host transformed by the recombinant vector described in Claim 4.

---

*Numbers in the margin indicate pagination in the foreign text.

(Claim 6) Method characterized by the fact that in manufacture of the serine protease described in Claim 1 or its partial peptide, the host described in Claim 5 is cultured and the abovementioned serine protease or its partial peptide is collected from the culture.

(Claim 7) Inhibitor screening method that uses the serine protease described in Claim 1 or its partial peptide.

Detailed Explanation of the Invention

(Industrial Field of Application)

This invention pertains to a novel serine protease, a gene that codes this, manufacture of said serine protease, and an inhibitor screening method that uses said serine protease.

(Prior Art)

Serine proteases are present widely in animals, plants, and microorganisms, and particularly in higher animals, are known to contribute to an extremely great number of biological reactions such as food digestion, blood coagulation and fibrinogenolysis, complement activation, hormone production, ovulation and insemination, phagocytosis, cell propagation, genesis and differentiation, aging, and cancer metastasis (Neurath, H. *Science*, **224**, 350-357, 1984). From the one-dimensional structure of their activation center, serine proteases in higher animals are classified as chymotrypsin and subtilisin types. It is known that in the chymotrypsin type, a histidine residual group in addition to a serine residual group in its activation center is essential to realize activation, and that amino acid sequences near the serine residual group and the histidine residual group are well-

preserved.

Therefore, cloning serine protease genes by the PCR method has been attempted using these preserved regions. That is, isolation of novel serine protease genes is reported using alanine-alanine-histidine-cysteine (AAHC) near the histidine residual group and aspartic acid-serine-glycine-glycine-proline (DSGGP) near the serine residual group, which are well-preserved in serine protease, as PCR primers.

For example, Sakanari et al. isolated a serine protease gene having 67% similarity to rat trypsin II from nematodes and protozoa (Sakanari, J. A., Staunton, C. E., Eakin, A. E., Carik, C. S., and McKerrow, J. H., *Proc. Natl. Acad. Sci. USA*, **86**, 4863-4867, 1989). In addition, the Mueller-Hill group isolated rat trypsin V and rat elastase IV from the rat pancreas (Kang, J., Wiegand, U., and Mueller-Hill, B., *Gene*, **110**, 181-187, 1992), and the same group also isolated trypsin IV from the human brain (Wiegand, U., Corbach, S. Minn, A. Kang, J., and Mueller-Hill, B., *Gene*, **136**, 167-175, 1993).

In these prior sources, however, genes are isolated based on nematodes and protozoa or cDNA derived from pancreas or brain tissue. In addition, because serine protease genes isolated using these types of PCR primers are present as zymogens, it is not confirmed at present whether or not they are genes coding proteins that have serine protease activity. Furthermore, it is not difficult to imagine that if genes could be propagated not only from nematodes, protozoa, and organs, but also from culture using cDNA of various types of implanted cancer cells, this would facilitate isolation of serine protease genes. However, at present, it is not possible to measure serine protease activity in the

supernatant of serum-added culture cells.

(Problems that the Invention is to Solve)

This invention was developed upon reflecting on the situation described above. Its purpose is to offer a novel serine protease and serine protease gene that codes this. A further purpose of this invention is to offer a method to mass-produce said protease using said gene, and a specific inhibitor screening method using said enzyme.

(Means of Solving the Problems)

The present inventors noted the human colon cancer cell COLO 201 as a departure material for isolating a novel serine protease gene. That is, the present inventors found serine protease enzyme activity in the supernatant of cell COLO 201 cultured in a non-protein culture, and discovered that using cDNA prepared from cancer cells such as cell COLO 201 was effective for isolating this type of novel serine protease gene. Moreover, to confirm whether or not the isolated gene truly is a gene that codes enzyme activity, they succeeded in manifesting this as a mature protein, and so perfected the present invention.

(Modes for Reducing the Invention to Practice)

The human colon cancer derived cell COLO 201 (ATCC CCL-224) can be cultured by any method normally used to culture animal cells. Moreover, it can be cultured by stationary culture using a culture medium that contains no protein. A concrete example is described in Working Example 1.

Enzyme activity in supernatant can be measured easily using a substance such as 7-amino-4-methylcumarin or p-nitroanilide bonded to a commercial synthetic substrate. A concrete example is described in

Working Example 2. As a result, clear serine protease enzyme activity was found in the culture supernatant of human colon cancer derived cell COLO 201. Therefore, the following test was conducted for the purpose of isolating all serine protease genes manifested in human colon cancer derived cell COLO 201 including this enzyme activity: mRNA was isolated and refined from human colon cancer derived cell COLO 201, and a cDNA library was fabricated. Cloning was performed by PCR using a PCR primer designed based on the serine protease motif from the fabricated cDNA library, and the PCR product obtained was subcloned.

As a result, a clone that includes a base sequence that codes the amino acids preserved in serine protease between the active residual groups serine and histidine was confirmed. As a result of cloning a full-length gene by standard method using the gene obtained in this way as a probe, gene SP59, gene SP60, and gene SP67 were isolated and novel serine protease could be confirmed. A concrete example is described in Working Example 3.

As a result of the above, the present inventors succeeded in isolating novel serine protease genes (gene SP59, gene SP60, and gene SP67) that have less than 30% similarity to existing serine protease from the cDNA of human colon cancer derived cell COLO 201. In addition, when manifestation of mRNA in human organs was confirmed using the isolated novel serine protease genes as probes, it was found that all of gene SP59, gene SP60, and gene SP67 were manifested in human organs, and gene SP59 showed especially strong manifestation in the brain at a size of approximately 1.4 kb. A concrete example is described in Working Example 4. From this fact, it was confirmed that the isolated serine

protease genes are manifested even in human organs.

In addition, from the structure of the isolated novel serine protease genes, a method was considered for manifesting these in animal cells as mature proteins. That is, it is known that by manifesting the typical serine protease of trypsin as the pro-form trypsinogen, then causing the enzyme enterokinase distributed in duodenal mucosa to act on this, it becomes present as a mature protein that has isoleucine as its N-terminal amino acid.

Therefore, a chimera gene (gene Trp59) was fabricated that connects the signal sequence of the trypsin gene in front of the gene considered to code the mature protein of gene SP59 to the gene that codes the enterokinase recognition sequence. The fabricated chimera gene of gene Trp59 was transfected to cell COS-1, then enterokinase was made to act on the culture supernatant of cell COS-1. As a result, serine protease enzyme activity was confirmed. A concrete example is described in Working Example 5.

From the above result, not only was it clear that serine protease genes isolated at this time were novel serine protease genes in terms of their primary structure; it was also clear that they manifested activity as mature proteins. In this invention, the nucleotide sequences of sequence numbers: 3, 4, and 5 are disclosed as nucleotide sequences of genes that code novel serine protease, but serine protease genes of this invention are not limited to these. Once the amino acid sequence of natural serine protease is determined, various nucleotide sequences that code the same amino acid sequence can be designed based on codon degeneration, and these can be prepared. In this case, it is preferred

to use a codon that is used with high frequency by the host to be used.

To obtain genes that code natural serine protease of this invention, cDNA can be obtained as described in Working Example 3, but this invention is not limited to this. That is, once one nucleotide sequence that codes the amino acid sequence of natural serine protease is determined, genes that code natural serine protease can be cloned as cDNA by different strategies from the strategies disclosed concretely in this invention, and furthermore, can be cloned from the genome of the cell that produces this.

When cloning from a genome, the various primer nucleotides or probe nucleotides used in Working Example 3 can be used as probes for selecting genome DNA fragments. In addition, other probes can be designed based on the nucleotide sequences described in sequence numbers: 3, 4, and 5. The general method for cloning an intended DNA from a genome is well-known in the art (*Current Protocols in Molecular Biology*, John Wiley & Sons, Chapters 5 and 6).

Genes that code natural serine protease of this invention can also be prepared by chemical synthesis. DNA can be chemically synthesized /4
easily by automatic DNA synthesizers used in the art; for example, by employing a synthesizer such as the 396 DNA/RNA synthesizer of Applied Biosystems. Therefore, DNA of the nucleotide sequences shown in sequence numbers: 3, 4, and 5 can be synthesized easily by persons skilled in the art.

Genes that code natural serine protease of this invention by a different codon from the biological codon can be prepared by chemical synthesis as described above, and DNA or RNA that has the nucleotide

sequences shown in sequence numbers: 3, 4, and 5 can be obtained following standard methods such as site-directed mutagenesis using a mutagenic primer as the matrix (see, for example, *Current Protocols in Molecular Biology*, John Wiley & Sons, Chapter 8).

When a serine protease gene of this invention is obtained as described above, this can be used to manufacture recombinant serine protease by standard genetic recombination. That is, DNA that codes serine protease of this invention is inserted into an appropriate manifest vector, said manifest vector is introduced into an appropriate host cell, said host cell is cultured, and the intended serine protease may be obtained from the culture obtained (cells or culture medium) in a biologically or chemically modified form; for example, N terminal acylation, examples of which are C_{1-6} acylation such as formylation or acetylation, or loss. The manifest system can also be designed to improve secretion efficiency and the amount manifested by adding or modifying the signal sequence or by the selection of host. An example of a means for adding or modifying the signal sequence is the method of linking a gene that codes the signal peptide of another structural peptide above the 5′ site of the structural gene of serine protease of this invention such that it is linked by way of a gene that codes a partial peptide that can be cut. A concrete example is the method described in Working Example 5 of using a gene that codes the signal sequence and enterokinase recognition sequence of the trypsin gene.

As hosts, protoskeletal organisms and true skeletal organisms can be used. Protoskeletal organisms that can be used include bacteria, especially *Escherichia coli* and *Bacillus* bacteria such as *B. subtilis*.

True skeletal organisms that can be used include yeasts, for example, *Saccharomyces* yeasts such as *S. serevisiae*, insect cells such as *Spodoptera frugiperda*, *Trichoplusia ni*, or *Bombyx mori*, and animal cells such as human cells, monkey cells, or mouse cells; concretely, cell COS-1, cell Vero, cell CHO, cell L, myocomma cells, cell C127, cell BALB/c3T3, or cell Sp-2/O. Furthermore, the organisms themselves can be used in this invention; for example, insects such as silkworms or cabbage loopers.

As manifest vectors, for example, plasmids, phages, phagemids, or viruses (Baculoviridae (insects) or vaccinia (animal)) can be used. The promoter in the manifest vector is selected depending on the host cell. For example, lac promotor or trp promotor are used as bacterial promotors, and adhl promotor or pqk promotor are used as yeast promotors. Examples of insect promotors include the Baculoviridae virus polypeptide lin promotor, and examples of animal promotors include Simian Virus 40 early or late promotor, CMV promotor, HSV-TK promotor, or Srα promotor. In addition, preferably, manifest vectors are used that besides the promotors described above, also contain elements such as enhancers, splicing signals, poly-A addition signals, and selection markers (for example, (methotrexate-resistant) dihydrofolic acid reducing enzyme gene or (G418-resistant) neo gene). Moreover, when using an enhancer, an enhancer such as SV40 enhancer is inserted above or below the gene.

Hosts can be transformed by a manifest vector by standard methods that are well-known in the art. These methods are described, for example, in *Current Protocols in Molecular Biology*, John Wiley & Sons.

In addition, the transformant can be cultured by standard methods. Serine protease can be refined from culture following standard methods such as limiting filtration or various types of column chromatography, such as chromatography using Sepharose.

Because the serine protease of this invention obtained in this way is a functional protein, this enzyme can be used to screen inhibitors specific to this enzyme, and said screening method is useful in research to search for drugs to treat various diseases. As a concrete example of a screening method, enzyme activity can be measured in the same way as in Working Example 2 for a test sample such as a peptide, protein, peptide-excluding compound, synthetic compound, or fermenter, or a natural component obtained from sources such as the supernatant of various types of cultures or an artificial component from sources such as various types of synthetic compounds. In addition, the screening method of this invention is a preferred mode for measuring enzyme activity as described above, or for other measurements such as bonding affinity measurement using a host or the cell wall part of a host that has been transformed either by DNA that codes a partial peptide of serine protease of this invention or by a gene of this enzyme described above or its partial peptide.

That is, serine protease of this invention can be used in the screening method of this invention in the form of its partial peptide. A host cell or cell wall part of a host cell transformed by a recombinant vector comprised by containing DNA that codes serine protease of this invention and manifests serine protease of this invention or its partial peptide may also be used in the screening

method of this invention.

Examples of such partial peptides include peptide fragments that are present near the serine residual group active site, and peptide fragments comprised of regions that have specificity to serine protease of this invention; for example, peptide fragments that can become recognition sites for antibodies having specificity to serine protease of this invention such as used in Working Example 3(6). Moreover, said partial peptides can be fabricated by the methods described above for serine protease of this invention or by already well-known peptide synthesis methods, or by cutting said serine protease by an appropriate protease.

The abovementioned "cell wall part" refers to a fraction containing many cell walls obtained after culturing a host cell that can manifest DNA that codes serine protease of this invention or its partial peptide under conditions that enable such manifestation, then pulverizing the host cells containing serine protease or its partial peptide obtained by an already well-known method.

The inhibitor screening method using serine protease of this invention or its partial peptide is performed by screening a test sample using serine protease of this invention or its partial peptide or a host cell or host cell wall parts that contain said serine protease of this invention or its partial peptide. A concrete example is screening by measuring enzyme activity or by measuring bonding affinity using a substrate of serine protease of this invention or its partial peptide; for example, a synthetic substrate such as a coloring substrate, or a substrate that has been marked by a radioactive species. Moreover, when

a host cell is used that contains serine protease, this can be used after fixing cells by an already well-known method (such as glutaraldehyde or formaldehyde).

(Working Examples)

Below, this invention is explained based on working examples.

Working Example 1. Preparation of Culture and Culture Supernatant of Human Colon Cancer Cell COLO 201

Human colon cancer cell COLO 201 (ATCC CCL-224) was cultured in a T flask (Nunc) that has a culture area of 80 cm^2. That is, 2×10^6 cells per T flask were implanted and cultured using RPMI-1640 culture (Nissui Seiyaku) containing 10% bovine fetal serum (FBS, GIBCO BRL Co.) until a confluent was formed. Next, this culture medium was replaced by RPMI-1640 that did not contain protein and contained 10^{-8} M sodium selenite (Sigma). After culturing for two weeks, the culture supernatant was collected, filtered and sterilized by a 0.22 μm sterilizing filter (Millipore), then supplied as a sample for measuring enzyme activity in the culture supernatant.

Working Example 2. Measurement of Enzyme Activity in Culture Supernatant of Human Colon Cancer Cell COLO 201

Serine protease activity in the culture supernatant obtained in Working Example 1 was measured using Test Team [as transliterated] coloring substance S-2251 (H-D-valeryl-L-leucyl-L-lysyl-p-nitroanilide dibasic salt, Daiichi Kagaku Yakuhin). That is, 50 μl Test Team coloring substance S-2251 dissolved 1 mg/ml in purified water, 40 μl 0.1 M Tris/HCl (pH 7.5), and 10 μl cell COLO 201 culture supernatant were combined and left 60 minutes at room temperature, then measured for

absorptance at 405 nm.

When absorptance after adding 10 μl culture instead of culture supernatant is taken as a blank, the absorptance of culture supernatant of cell COLO 201 was 0.42. In addition, this showed comparable activity even using H-D-valeryl-L-leucyl-L-arginyl-p-nitroanilide dibasic salt (Daiichi Kagaku Yakuhin). As a result of considering the effect of various types of protease inhibitors in this measurement system, it was confirmed that culture supernatant of cell COLO 201 had clear serine protease enzyme activity (Table 1).

TABLE 1 /6

Inhibitor* or Treatment		Surviving Activity (%)
aprotinin	250 KIU/ml	0.4%
leupeptin	0.1 mM	0.7%
benzamidine	1 mM	0.7%
pABSF[1]	1 mM	1.4%
NEM[2]	1 mM	100.0%
EDTA[3]	1 mM	74.0%
triton	2.5%	61.1%
	0.25%	100.0%
SDS[4]	0.2%	0.0%
heating	95°C, 10 min	27.0%

* pre-incubation: 37°C, 10 min

1. pABSF: 4-(2-aminoethyl)-benzenesulfonyl fluoride · HCl (Wako Pure Chemicals)
2. NEM: N-ethylmaleimide
3. EDTA: ethylenediamine tetraacetic acid (Sigma)
4. SDS: sodium dodecylsulfate (Sigma)

Working Example 3. Cloning of Novel Serine Protease Gene and Identification of Protein

(1) Isolation and Refining of mRNA of Cell COLO 201

mRNA of cell COLO 201 was prepared using Isogen (Nippon Gene) according to the appended documentation. That is, cell COLO 201 was propagated in a T flask (Nunc, 80 cm^2) until a confluent was formed, then cells were diluted by adding 1 ml Isogen per T flask. Furthermore, this was combined with 200 μl chloroform and agitated, then centrifuged 15 minutes at 15,000 rpm and 4°C.

After centrifuging, the water phase was collected. The collected water phase was combined with 500 μl isopropanol and agitated, then centrifuged 30 minutes at 15,000 rpm and 4°C. All of the RNA sediment obtained was dissolved in 400 μl distilled water treated with diethyl procarbonate (DEPC), and combined and mixed with 400 μl 2× elution buffer (20 mM Tris-HCl pH 7.5, 2 mM EDTA, 0.2% SDS). Furthermore, this was combined and mixed with 500 μl Oligotex-dT30 (NipponRoche) suspension, and heated 5 minutes at 65°C. After cooling in ice water, this was combined with 130 μl 5 M NaCl and heated 10 minutes at 37°C.

After heating, this was centrifuged 3 minutes at 15,000 rpm and 4°C, the supernatant was removed, then the sediment was suspended in 500 μl washing buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% SDS, 0.1 M NaCl) and centrifuged 3 minutes more at 15,000 rpm and 4°C. After again removing the supernatant, the sediment was suspended in 400 μl DEPC-treated distilled water. This was heated 5 minutes at 65°C, then centrifuged 3 minutes more at 15,000 rpm and 4°C, and the supernatant was collected.

This supernatant was combined with 20 μl 5 M NaCl and 1 ml ethanol and agitated, then centrifuged 20 minutes more at 15,000 rpm and 4°C. The sediment was washed in 500 μl 70% ethanol and lightly air-dried, then dissolved in 10 μl DEPC-treated distilled water. As a result, approximately 12 μg polyA$^+$ RNA were obtained from 16 T flasks.

(2) Preparation of cDNA Library

A cDNA library was prepared using the Super Script Plasmid System (Life Technologies).

Step 1. Synthesis of cDNA

5 μl (approximately 6 μg) cell COLO 201 mRNA were combined with 2 μl (1 μg) Oligo dT NotI primer and heated 10 minutes at 70°C, then cooled in ice water. This heat-modified mRNA was combined with 4 μl 5× first strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$), 1 μl 10 mM dNTP, 2 μl 0.1 M DTT, DEPC-treated distilled water, and 5 μl (1000 U) Super Script II RT, and reacted 1 hour at 37°C.

Next, this reaction solution was combined with 91 μl DEPC-treated distilled water, 30 μl 5× second strand buffer (100 mM Tris-HCL pH 6.9, 450 mM KCl, 23 mM $MgCl_2$, 0.75 mM β-NAD$^+$, 50 mM $(NH_4)_2SO_4$), 3 μl 10 mM dNTP, 1 μl (10 U) *E. coli* DNA ligase, 4 μl (40 U) *E. coli* DNA polymerase, and 1 μl (2 U) *E. coli* RNase H and heated 2 hours at 16°C, then 2 μl (10 U) T4 DNA polymerase were added and reacted 5 minutes at 16°C. /7

Furthermore, this solution was combined and mixed with 10 μl 0.5 M EDTA, then combined with 150 μl phenol:chloroform:isoamyl alcohol (25:24:1). This was agitated, then centrifuged 5 minutes at 15,000 rpm, and the supernatant was collected. The supernatant collected was combined with 10 μl 5 M KOAc and 400 μl ethanol, agitated, and

centrifuged 10 minutes at 15,000 rpm. The sediment obtained by centrifuging was washed in 500 μl 70% ethanol and lightly air-dried, then dissolved in 25 μl DEPC-treated distilled water.

Step 2. Addition of Sal I Adapter

25 μl two-chain cDNA obtained in Step 1 were combined with 10 μl 5× T4 DNA ligase buffer (250 mM Tris-HCl pH 7.6, 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25% (w/v), PEG 8000), 10 μl (10 μg) Sal I adapter solution, and 5 μl (5 U) T4 DNA ligase and reacted 16 hours at 16°C, then combined with 50 μl phenol:chloroform:isoamyl alcohol (25:24:1). This was agitated, then centrifuged 5 minutes at 15,000 rpm, and the supernatant was collected. The supernatant collected was combined with 5 μl 5 M KOAc and 125 μl ethanol, agitated, cooled 20 minutes at -80°C, and centrifuged 10 minutes at 15,000 rpm. The sediment obtained by centrifuging was washed in 200 μl 70% ethanol and lightly air-dried, then dissolved in 40 μl DEPC-treated distilled water.

Step 3. Cutting by Restriction Enzyme Not I

20 μl reaction solution of Step 2 were combined with 4 μl (60 U) Not I and reacted 3 hours at 37°C, then was extracted by phenol:chloroform:isoamyl alcohol (25:24:1) and the supernatant was collected. This supernatant was fractionated to a size of 1 kilo base pairs or greater by a Chromaspin-1000 column (Chrontek), and 50 μl eluate were obtained.

Step 4. Ligation with pSPORT Vector

3 μl size-fractionated cDNA solution were combined with 1 μl pSPORT vector (50 ng; Life Technologies) consumed by Sal I and Not I, then further combined with 11 μl DEPC-treated distilled water, 4 μl 5× T4 DNA

ligase buffer, and 1 μl 5× T4 DNA ligase and reacted 3 hours at room temperature.

After reacting, this was extracted by phenol:chloroform:isoamyl alcohol (25:24:1), and 5 μl (5 μg) yeast tRNA, 5 μl 5 M KOAc, and 125 μl ethanol were added. This was agitated and cooled 20 minutes at -80°C, then centrifuged 10 minutes at 15,000 rpm. The sediment obtained by centrifuging was washed in 200 μl 70% ethanol and lightly air-dried, then dissolved in 5 μl TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

Step 5. Transformation to *E. coli* DH10B

The ligated cDNA obtained in Step 4 was transformed by the electroporation method to *E. coli* Electro MAX DH10B (F′, mcrA, φ 80dlacZΔM15, Δ(mrr-hsdRMS-mcrBC), ΔlacX74, deoR, recA1, endA1, araD139, Δ(ara, leu)7697, galU, galK, λ-, rpsL, nupG: Life Technologies). That is, 50 μl cell DH10B were combined with 2 μl ligated cDNA to a final volume of 26 μl × 2, then treated by an electroporator (Bio-Rad) under conditions of 400 V and 330 μF.

Next, *E. coli* was collected in 4 ml SOC culture (2% packed trypsin, 0.5% packed yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 20 mM glucose) and cultured 1 hour by shaking culture at 37°C, then wrapped in an LB plate containing 50 mg/ml ampicillin (1% packed trypsin, 0.5% packed yeast extract, 0.5% NaCl, 0.1% glucose, 1.5% packed agar) and cultured overnight at 37°C. As a result, a cDNA library containing approximately 1.1×10^6 clones was obtained.

(3) PCR Using Serine Protease Preserved Regions

Oligomer KY185 that shows sequence number: 1 based on the amino acid preserved region near active residual group (His) was synthesized.

In addition, oligomer KY189 that shows sequence number: 2 based on the amino acid preserved region near active residual group (Ser) was synthesized. PCR was performed by Ampli-Taq polymerase (Parkin Elmer Co.) with the cDNA obtained in Working Example 3(2) Step 3 as template and oligomers KY185 and KY189 as primers. This PCR reaction solution was subcloned by pCR II vector (Invitrogen), and clones were obtained that have a DNA fragment with 431 base pairs. As a result of sequencing these clones, it was confirmed that they contained a base sequence that codes the amino acid sequence preserved in serine protease between two active residual groups (His) and (Ser).

(4) Sequencing of Serine Protease

A fluorescent-marked probe was fabricated by PCR using the plasmid obtained in Working Example 3(3) described above as template. Using this probe, the cDNA library of approximately 1,100,000 clones obtained in Working Example 3(2) Step 5 was screened by standard method. As a result, from approximately 200,000 clones, six positive clones were obtained. The size of the inserted DNA fragment was studied, the longest clone pSPORT/SP59-#3 (approximately 1.4 kilo base pairs) was selected, and the sequence of this gene was determined by a Taq Dye Deoxy Terminator Cycle Sequencing Kit (Applied Biosystems).

(5) Base Sequence Characteristics

The cDNA base sequence of pSPORT/SP59-#3 is shown by sequence number: 3. As a result, the cDNA of pSPORT/SP59-#3 has a total length of 1,438 base pairs, and is comprised of the 5′ nontranslation region of base pair 155, the translation region of base pair 732, and the 3′ nontranslation region of base pair 551. It was clear that the

translation region codes the amino acid 244 residual group.

(6) Fabrication of Antibody to Peptide Fragment of Protein SP59 /8

Of the amino acid sequence of SP59, a partial peptide with sequence number: 6 (Cys added to amino acid numbers 56 to 67 of sequence number: 3), a partial peptide with sequence number: 7 (amino acid numbers 96 to 110 of sequence number: 3), and a partial peptide with sequence number: 8 (amino acid numbers 210 to 223 of sequence number: 3) were synthesized. Each partial peptide was obtained at a purity of 90% or greater.

Each partial peptide was inoculated by bonding with bovine serum albumin (BSA, Nakaraitesk [as transliterated]) activated by N-(m-maleimidobenzoyloxy)succinimide (MBS, Nakaraitesk). That is, 5 mg BSA were dissolved in 50 mM phosphoric acid buffer (pH 8.0), then 1.25 mg MBS dissolved in DMSO were added and agitated 30 minutes at room temperature, and MBS-activated BSA was obtained. Next, 5 mg of each partial peptide dissolved in 50 mM phosphoric acid buffer (pH 7.0) were added to MBS-activated BSA and coupled by agitating 3 hours at room temperature. Each of the coupled partial peptides was mixed with Freund's complete adjuvant (Nakaraitesk), and antiserum was prepared by standard method.

(7) Refining of Protein SP59 from Culture Supernatant of Human Pancreatic Cancer Cell HPC-Y3

10 mg freeze-dried product of culture supernatant of cell HPC-Y3 obtained in the same way as in Working Example 1 were dissolved in 1 mg/ml 10 mM Tris/HCl pH 7.4 containing 0.1 M NaCl and supplied to gel filtration chromatography using Superose 6 (Pharmacia) at a flow speed

of 4 ml/min. Each fraction was blotted by Weston blot using the SP59 partial peptide antibody obtained by (6), and was measured for enzyme activity using synthetic matrices (Boc-Phe-Ser-Arg-4-methyl-cumaryl-7-amide (hereinafter called MCA) and Boc-Gln-Ala-Arg-MCA). As a result, activity was found in the fragment eluted as fraction 63-70. This fraction was applied as is to ion exchange chromatography by a MonoQ column (Pharmacia).

Next, when the fraction not bonded to the MonoQ column was applied as is to a hydroxyapatite column (Pentax) buffered ahead of time by 10 mM phosphoric buffer pH 6.8, then eluted by a linear gradient of phosphoric buffer, the active fraction eluted by phosphoric buffer had a concentration of 150 mM. Next, this was applied to a MonoS column buffered ahead of time by 20 mM phosphoric buffer pH 6.8, and an active fraction was eluted that had a single peak at 0.1 M NaCl concentration. The eluted fraction was desalinated by a C4 column, then supplied to N-terminal amino acid analysis.

(8) Analysis of N-Terminal Amino Acid

N-terminal amino acid analysis of protein SP59 was performed as follows: SDS-polyacrylamide electrophoresis was performed from the non-protein culture supernatant of cell HPC-Y3 using protein SP59 refined by the method described above. After electrophoresis, this was transferred to PVDF film following the method of Matsudaira (Matsudaira, P. (1987) *J. Biol. Chem.*, **262**, 10035-10038). Furthermore, protein SP59 was detected by Coomassie blue staining following the method of Speicher (Speicher, D. W. (1989), *Techniques in Protein Chemistry* (Hugli, T. E., ed.) pp. 24-35, Academic Press, San Diego). This stained protein SP59

fragment was cut, washed well and dried, and supplied as a sample for N-terminal amino acid analysis. An Applied Biosystems 477A gas phase sequencer was used for this analysis.

The phenylthiohydantoin derivative was identified by Applied Biosystems 120A on-line system reverse-phase HPLC (Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981), *J. Biol. Chem.*, **256**, 7990-7997). As a result, it was confirmed that, as surmised, the mature N-terminal amino acid sequence of protein SP59 was the amino acid sequence (LVHG). In addition, it was clear that amino acid sequence (VHG) lacking the one N-terminal amino acid of leucine of protein SP59 was present simultaneously.

(9) Cloning and Protein Identification of Genes SP60 and SP67

Genes SP60 and SP67 were cloned from cell COLO 201 and their proteins identified in the same way as in the method described above, and SP60 (sequence number: 4) and SP67 (sequence number: 5) were obtained that have a catalytic triad residual group that has specificity to serine protease. DNA of these can be manifested and serine protease can be obtained in the same way as SP59.

Working Example 4. Manifestation of Gene SP59 in Human Organs by Northern Blotting

pSPORT/SP59-#3 was consumed by restriction enzyme Mlu I, a DNA fragment of approximately 1.4 kilo base pairs was isolated and refined, and this was marked by α-^{32}P dCTP (Amersham) and made a probe. This probe and a membrane filter (Chrontek) blotted with mRNA prepared from 16 types of organs were reacted 2 hours at 65°C.

Next, this membrane filter was washed twice, once for 20 minutes at

room temperature in 2× SSC (150 mM NaCl, 15 mM sodium ascorbate) containing 0.1% SDS, then for 30 minutes at 65°C replacing this with 1× SSC and 0.1% SDS. Next, this was exposed for 30 minutes to an imaging plate for BAS2000 (Fuji Photo Film) and analyzed. Results are shown in Figure 1. Manifestation of mRNA of SP59 in human organs was found to be especially strong in the brain, at a size of approximately 1.4 kb. In addition, as a result of testing gene SP60 and gene SP67 in the same way as gene SP59, it was found that gene SP60 was manifested strongly in the colon, prostate, and kidney, and SP67 was manifested strongly in the colon, small intestine, prostate, and pancreas.

Working Example 5. Measurement of Enzyme Activity of Novel Serine Protease Mature Protein that Codes Gene SP59

(1) Construction of Manifest Plasmid

pSPORT/SP59-#3 was consumed by restriction enzyme Mlu I, then a DNA fragment of approximately 1.4 kilo base pairs was isolated, refined, and dissolved in TE. Similarly, pdKCR vector that has an SV40 promotor (Nikaido, T. et al., *Nature*, **311**, 631-635 (1984): vector that has pBR327 substituted in the pBR322 site of pKCR vector) was consumed by Mlu I, then was dephosphated by alkali phosphatase, extracted by phenol:chloroform:isoamyl alcohol (25:24:1), precipitated in ethanol, and dissolved in TE.

pSPORT/SP59-#3 DNA fragment and pdKCR vector DNA fragment were ligated following standard method, *E. coli* JM109 was transformed, and the colony produced by the PCR method was analyzed to obtain manifest plasmid pdKCR/SP59 of the intended serine protease SP59. Next, a gene was propagated that codes the signal sequence following initial

methionine and the enterokinase recognition sequence of trypsin II, and primers were designed such that an Eco RI restriction enzyme site was added above 5′ and a Bsp MI restriction enzyme site was added above 3′. KY239 and KY240 are shown by sequence number: 9 and sequence number: 10.

Using these primers KY239 and KY240, PCR was performed with pCR II/Trypsin II plasmid as template (obtained by propagating by the cDNA library obtained by Working Example 3(2) Step 5 using two specific primers (Emi, M., Nakamura et al., *Gene*, **41**, 305-310, 1986), then subcloning by pCRII vector). After consuming the product by restriction enzymes (Eco RI and Bsp MI), a DNA fragment of approximately 75 bp was isolated and refined.

Similarly, primers KY241 and KY207 were designed such that a Bsp MI restriction enzyme recognition site was added above the gene that codes the mature protein of gene SP59. KY241 and KY207 are shown by sequence number: 11 and 12. Using these primers KY241 and KY207, PCR was performed with pSPORT/SP59-#3 plasmid as template. After consuming the product by restriction enzymes (Bsp MI and Bpu 1102I), a DNA fragment was isolated and refined. Next, the DNA fragment that codes the signal sequence and enterokinase recognition sequence of trypsin II obtained and the DNA fragment that codes the mature protein of gene SP59 were ligated to a pdKCR/SP59 vector pre-consumed by restriction enzymes (Eco RI and Bpu 1102I) following standard method, and *E. coli* JM109 was transformed. From the transformed colony, a colony that contains the intended chimera gene was confirmed by the PCR method, and a manifest plasmid (pdKCR/Trp59) of the intended chimera gene (Trp59) was obtained.

(2) Manifestation in Cell COS-1

Manifestation in animal cells was attempted using the manifest plasmid of the intended chimera gene (Trp59) obtained in Working Example 5(1). Using cell COS-1 as the animal cell for manifestation, this was transfected by the lipofectin method with each of pdKCR/Trp59 and pdKCR as manifest plasmids. That is, 1×10^6 cells of cell COS-1 were implanted in a culture dish with a diameter of 10 cm (Corning, 430167). Dulbecco's minimum essential medium (DMEM, Nissui Seiyaku) containing 10% bovine fetal serum was used as the culture medium.

The next day, cells were rinsed by 5 ml Opti-MEM culture (Life Technologies), then combined with 5 ml more Opti-MEM culture and cultured 2 hours at 37°C. After culturing, a mixture of 1 μg of the plasmid described above and 10 μg lipofectin (Pharmacia) were added per dish, and cells were cultured 5 hours more at 37°C. After culturing, 5 ml Opti-MEM culture were added for a total of 10 ml, and cells were cultured 72 hours at 37°C. After culturing, the culture supernatant was collected by centrifugation and used as a sample for measuring enzyme activity.

(3) Measurement of Enzyme Activity

Enzyme activity in the culture supernatant obtained in Working Example 5(2) was measured. That is, 10 μl enterokinase (1 mg/ml, Biozyme Laboratories) were mixed with 50 μl culture supernatant of cell COS-1 and reacted 15 minutes at room temperature. Next, this was combined with 50 μl 0.2 M matrix solution of synthetic matrix Boc-Phe-Ser-Arg-MCA dissolved in DMSO (Peptide Laboratories) diluted with 0.1 M Tris/HCl pH 8.0, and reacted 60 minutes more at room temperature. After reacting,

fluorescence was measured at an excitation wavelength of 485 nm and a fluorescent wavelength of 535 nm.

As shown in Figure 2, results confirmed enzyme activity by adding enterokinase to culture supernatant of cell COS-1 in which gene Trp59 was manifested. As a result, it was clear that the novel serine protease mature protein that codes gene SP59 shows enzyme activity. From the above result, not only is it clear that the serine protease gene isolated at this time is a novel serine protease gene in terms of its primary structure; it is also clear that it manifests activity as a mature protein.

(Effects of the Invention)

The present inventors isolated a novel serine protease gene from human colon cancer derived cell COLO 201, and moreover, demonstrated that the isolated gene has enzyme activity. In addition, they demonstrated that in spite of the fact that the novel serine protease gene first obtained at this time is derived from colon cancer, gene SP59 is manifested strongly in the human brain.

Thus, this clearly shows that isolation of serine protease genes using cancel cells is useful as a novel gene resource. Furthermore, even when isolating a novel serine protease gene, or even when studying manifestation of mRNA using the isolated gene, there is no guarantee that the translated protein will be functionally manifested in its organ site.

The fact that it was clear that a novel serine protease gene is manifested by the method described above and that it codes a functional protein proves the usefulness of said gene. In addition, because the

protein manifested using said gene is a functional protein, by establishing a screening system for inhibitors specific to this enzyme, it becomes possible for the first time to screen drugs to treat various diseases.

(Sequence Charts)

Sequence number: 1

Sequence number: 20

Type of sequence: nucleic acid

Number of chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

GTGCTCACNG CNGCBCAYTG 20

Sequence number: 2

Sequence length: 20

Type of sequence: nucleic acid

Number of chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

AGCGGNCCNC CDGARTCVCC 20

Sequence number: 3

Sequence length: 1438

Type of sequence: nucleic acid

Number of chains: two chains

Topology: straight-chain

Class of sequence: cDNA to mRNA

Sequence

```
GGACACACGC TGTAGCTGTC TCCCCGGCTG GCTGGCTCGC TCTCTCCTGG GGACACAGAG   60
GTCGGCAGGC AGCACACAGA GGGACCTACG GGCAGCTGTT CCTTCCCCCG ACTCAAGAAT  120
                                  20
CCCCGGAGGC CCGGAGGCCT GCAGCAGGAG CGGCC ATG AAG AAG CTG ATG GTG      173
                                       Met Lys Lys Leu Met Val
                                         -20
GTG CTG AGT CTG ATT GCT GCA GCC TGG GCA GAG GAG CAG AAT AAG TTG      221
Val Leu Ser Leu Ile Ala Ala Ala Trp Ala Glu Glu Gln Asn Lys Leu
-15                 -10                 -5                  -1  1
GTG CAT GGC GGA CCC TGC GAC AAG ACA TCT CAC CCC TAC CAA GCT GCC      269
Val His Gly Gly Pro Cys Asp Lys Thr Ser His Pro Tyr Gln Ala Ala
             5                  10                  15
CTC TAC ACC TCG GGC CAC TTG CTC TGT GGT GGG GTC CTT ATC CAT CCA      317
Leu Tyr Thr Ser Gly His Leu Leu Cys Gly Gly Val Leu Ile His Pro
        20                  25                  30
CTG TGG GTC CTC ACA GCT GCC CAC TGC AAA AAA CCG AAT CTT CAG GTC      365
Leu Trp Val Leu Thr Ala Ala His Cys Lys Lys Pro Asn Leu Gln Val
    35                  40                  45
TTC CTG GGG AAG CAT AAC CTT CGG CAA AGG GAG AGT TCC CAG GAG CAG      413
Phe Leu Gly Lys His Asn Leu Arg Gln Arg Glu Ser Ser Gln Glu Gln
 50                  55                  60                  65
AGT TCT GTT GTC CGG GCT GTG ATC CAC CCT GAC TAT GAT GCC GCC AGC      461
Ser Ser Val Val Arg Ala Val Ile His Pro Asp Tyr Asp Ala Ala Ser
                70                  75                  80
CAT GAC CAG GAC ATC ATG CTG TTG CGC CTG GCA CGC CCA GCC AAA CTC      509
His Asp Gln Asp Ile Met Leu Leu Arg Leu Ala Arg Pro Ala Lys Leu
            85                  90                  95
TCT GAA CTC ATC CAG CCC CTT CCC CTG GAG AGG GAC TGC TCA GCC AAC      557
Ser Glu Leu Ile Gln Pro Leu Pro Leu Glu Arg Asp Cys Ser Ala Asn
        100                 105                 110

ACC ACC AGC TGC CAC ATC CTG GGC TGG GGC AAG ACA GCA GAT GGT GAT      605
Thr Thr Ser Cys His Ile Leu Gly Trp Gly Lys Thr Ala Asp Gly Asp
```

```
    115                 120                 125
TTC CCT GAC ACC ATC CAG TGT GCA TAC ATC CAC CTG GTG TCC CGT GAG    653
Phe Pro Asp Thr Ile Gln Cys Ala Tyr Ile His Leu Val Ser Arg Glu
130                 135                 140                 145
GAG TGT GAG CAT GCC TAC CCT GGC CAG ATC ACC CAG AAC ATG TTG TGT    701
Glu Cys Glu His Ala Tyr Pro Gly Gln Ile Thr Gln Asn Met Leu Cys
                150                 155                 160
GCT GGG GAT GAG AAG TAC GGG AAG GAT TCC TGC CAG GGT GAT TCT GGG    749
Ala Gly Asp Glu Lys Tyr Gly Lys Asp Ser Cys Gln Gly Asp Ser Gly
            165                 170                 175
GGT CCG CTG GTA TGT GGA GAC CAC CTC CGA GGC CTT GTG TCA TGG GGT    797
Gly Pro Leu Val Cys Gly Asp His Leu Arg Gly Leu Val Ser Trp Gly
        180                 185                 190
AAC ATC CCC TGT GGA TCA AAG GAG AAG CCA GGA GTC TAC ACC AAC GTC    845
Asn Ile Pro Cys Gly Ser Lys Glu Lys Pro Gly Val Tyr Thr Asn Val
    195                 200                 205
TGC AGA TAC ACG AAC TGG ATC CAA AAA ACC ATT CAG GCC AAG            887
Cys Arg Tyr Thr Asn Trp Ile Gln Lys Thr Ile Gln Ala Lys
210                 215                 220
TGACCCTGAC ATGTGACATC TACCTCCCGA CCTACCACCC CACTGGCTGG TTCCAGAACG  947
TCTCTCACCT AGACCTTGCC TCCCCTCCTC TCCTGCCCAG CTCTGACCCT GATGCTTAAT 1007
AAACGCAGCG ACGTGAGGGT CCTGATTCTC CCTGGTTTTA CCCCAGCTCC ATCCTTGCAT 1067
CACTGGGGAG GACGTGATGA GTGAGGACTT GGGTCCTCGG TCTTACCCCC ACCACTAAGA 1127
GAATACAGGA AAATCCCTTC TAGGCATCTC CTCTCCCCAA CCCTTCCACA CGTTTGATTT 1187
CTTCCTGCAG AGGCCCAGCC ACGTGTCTGG AATCCCAGCT CCGCTGCTTA CTGTCGGTGT 1247
CCCCTTGGGA TGTACCTTTC TTCACTGCAG ATTTCTCACC TGTAAGATGA AGATAAGGAT 1307
GATACAGTCT CCATAAGGCA GTGGCTGTTG GAAAGATTTA AGGTTTCACA CCTATGACAT 1367
ACATGGAATA GCACCTGGGC CACCATGCAC TCAATAAAGA ATGAATTTTA TTAAAAAAAA 1427
AAAAAAAAAA A                                                      1438
```

Sequence number: 4

Sequence length: 699

Type of sequence: nucleic acid

Number of chains: two chains

Topology: straight-chain

Class of sequence: cDNA to mRNA

Sequence

```
GTG GTG GGT GGG GAG GAG GCC TCT GTG GAT TCT TGG CCT TGG CAG GTC     48
Val Val Gly Gly Glu Glu Ala Ser Val Asp Ser Trp Pro Trp Gln Val
 1               5                  10                  15
AGC ATC CAG TAC GAC AAA CAG CAC GTC TGT GGA GGG AGC ATC CTG GAC     96
Ser Ile Gln Tyr Asp Lys Gln His Val Cys Gly Gly Ser Ile Leu Asp
            20                  25                  30
CCC CAC TGG GTC CTC ACG GCA GCC CAC TGC TTC AGG AAA CAT ACC GAT    144
Pro His Trp Val Leu Thr Ala Ala His Cys Phe Arg Lys His Thr Asp
        35                  40                  45
GTG TTC AAC TGG AAG GTG CGG GCA GGC TCA GAC AAA CTG GGC AGC TTC    192
Val Phe Asn Trp Lys Val Arg Ala Gly Ser Asp Lys Leu Gly Ser Phe
    50                  55                  60
CCA TCC CTG GCT GTG GCC AAG ATC ATC ATC ATT GAA TTC AAC CCC ATG    240
Pro Ser Leu Ala Val Ala Lys Ile Ile Ile Ile Glu Phe Asn Pro Met
65                  70                  75                  80
TAC CCC AAA GAC AAT GAC ATC GCC CTC ATG AAG CTG CAG TTC CCA CTC    288
Tyr Pro Lys Asp Asn Asp Ile Ala Leu Met Lys Leu Gln Phe Pro Leu
```

```
                    85                  90                  95

ACT TTC TCA GGC ACA GTC AGG CCC ATC TGT CTG CCC TTC TTT GAT GAG   336
Thr Phe Ser Gly Thr Val Arg Pro Ile Cys Leu Pro Phe Phe Asp Glu
            100                 105                 110
GAG CTC ACT CCA GCC ACC CCA CTC TGG ATC ATT GGA TGG GGC TTT ACG   384
Glu Leu Thr Pro Ala Thr Pro Leu Trp Ile Ile Gly Trp Gly Phe Thr
        115                 120                 125
AAG CAG AAT GGA GGG AAG ATG TCT GAC ATA CTG CTG CAG GCG TCA GTC   432
Lys Gln Asn Gly Gly Lys Met Ser Asp Ile Leu Leu Gln Ala Ser Val
    130                 135                 140
CAG GTC ATT GAC AGC ACA CGG TGC AAT GCA GAC GAT GCG TAC CAG GGG   480
Gln Val Ile Asp Ser Thr Arg Cys Asn Ala Asp Asp Ala Tyr Gln Gly
145                 150                 155                 160
GAA GTC ACC GAG AAG ATG ATG TGT GCA GGC ATC CCG GAA GGG GGT GTG   528
Glu Val Thr Glu Lys Met Met Cys Ala Gly Ile Pro Glu Gly Gly Val
                165                 170                 175
GAC ACC TGC CAG GGT GAC AGT GGT GGG CCC CTG ATG TAC CAA TCT GAC   576
Asp Thr Cys Gln Gly Asp Ser Gly Gly Pro Leu Met Tyr Gln Ser Asp
            180                 185                 190
CAG TGG CAT GTG GTG GGC ATC GTT AGC TGG GGC TAT GGC TGC GGG GGC   624
Gln Trp His Val Val Gly Ile Val Ser Trp Gly Tyr Gly Cys Gly Gly
        195                 200                 205
CCG AGC ACC CCA GGA GTA TAC ACC AAG GTC TCA GCC TAT CTC AAC TGG   672
Pro Ser Thr Pro Gly Val Tyr Thr Lys Val Ser Ala Tyr Leu Asn Trp
    210                 215                 220
ATC TAC AAT GTC TGG AAG GCT GAG CTG                              699
Ile Tyr Asn Val Trp Lys Ala Glu Leu
225                 230
```

Sequence number: 5

Sequence length: 723

Type of sequence: nucleic acid

Number of chains: two chains

Topology: straight-chain

Class of sequence: cDNA to mRNA

Sequence

```
GTT GTT GGG GGC ACG GAT GCG GAT GAG GGC GAG TGG CCC TGG CAG GTA     48
Val Val Gly Gly Thr Asp Ala Asp Glu Gly Glu Trp Pro Trp Gln Val
  1               5                  10                  15
AGC CTG CAT GCT CTG GGC CAG GGC CAC ATC TGC GGT GCT TCC CTC ATC     96
Ser Leu His Ala Leu Gly Gln Gly His Ile Cys Gly Ala Ser Leu Ile
            20                  25                  30
TCT CCC AAC TGG CTG GTC TCT GCC GCA CAC TGC TAC ATC GAT GAC AGA    144
Ser Pro Asn Trp Leu Val Ser Ala Ala His Cys Tyr Ile Asp Asp Arg
        35                  40                  45
GGA TTC AGG TAC TCA GAC CCC ACG CAG TGG ACG GTC TTC CTG GGC TTG    192
Gly Phe Arg Tyr Ser Asp Pro Thr Gln Trp Thr Val Phe Leu Gly Leu
    50                  55                  60
CAC GAC CAG AGC CAG CGC AGC GCC CCT GGG GTG CAG GAG CGC AGG CTC    240
His Asp Gln Ser Gln Arg Ser Ala Pro Gly Val Gln Glu Arg Arg Leu
65                  70                  75                  80
AAG CGC ATC ATC TCC CAC CCC TTC TTC AAT GAC TTC ACC TTC GAC TAT    288
Lys Arg Ile Ile Ser His Pro Phe Phe Asn Asp Phe Thr Phe Asp Tyr
```

```
                    85                  90                  95
GAC ATC GCG CTG CTG GAG CTG GAG AAA CCG GCA GAG TAC AGC TCC ATG    336
Asp Ile Ala Leu Leu Glu Leu Glu Lys Pro Ala Glu Tyr Ser Ser Met
            100                 105                 110

GTG CGG CCC ATC TGC CTG CCG GAC GCC TCC CAT GTC TTC CCT GCC GGC    384
Val Arg Pro Ile Cys Leu Pro Asp Ala Ser His Val Phe Pro Ala Gly
        115                 120                 125
AAG GCC ATC TGG GTC ACG GGC TGG GGA CAC ACC CAG TAT GGA GGC ACT    432
Lys Ala Ile Trp Val Thr Gly Trp Gly His Thr Gln Tyr Gly Gly Thr
    130                 135                 140
GGC GCG CTG ATC CTG CAA AAG GGT GAG ATC CGC GTC ATC AAC CAG ACC    480
Gly Ala Leu Ile Leu Gln Lys Gly Glu Ile Arg Val Ile Asn Gln Thr
145                 150                 155                 160
ACC TGC GAG AAC CTC CTG CCG CAG CAG ATC ACG CCG CGC ATG ATG TGC    528
Thr Cys Glu Asn Leu Leu Pro Gln Gln Ile Thr Pro Arg Met Met Cys
                165                 170                 175
GTG GGC TTC CTC AGC GGC GGC GTG GAC TCC TGC CAG GGT GAT TCC GGG    576
Val Gly Phe Leu Ser Gly Gly Val Asp Ser Cys Gln Gly Asp Ser Gly
            180                 185                 190
GGA CCC CTG TCC AGC GTG GAG GCG GAT GGG CGG ATC TTC CAG GCC GGT    624
Gly Pro Leu Ser Ser Val Glu Ala Asp Gly Arg Ile Phe Gln Ala Gly
        195                 200                 205
GTG GTG AGC TGG GGA GAC GGC TGC GCT CAG AGG AAC AAG CCA GGC GTG    672
Val Val Ser Trp Gly Asp Gly Cys Ala Gln Arg Asn Lys Pro Gly Val
    210                 215                 220
TAC ACA AGG CTC CCT CTG TTT CGG GAC TGG ATC AAA GAG AAC ACT GGG    720
Tyr Thr Arg Leu Pro Leu Phe Arg Asp Trp Ile Lys Glu Asn Thr Gly
225                 230                 235                 240
GTA                                                            723
Val
```

Sequence number: 6

Sequence length: 13

Type of sequence: amino acid

Topology: straight-chain

Class of sequence:

Sequence

Leu Arg Gln Arg Glu Ser Ser Gln Glu Gln Ser Ser Cys
1 5 10

Sequence number: 7

Sequence length: 15

Type of sequence: amino acid

Topology: straight-chain

Class of sequence:

Sequence

Lys Leu Ser Glu Leu Ile Gln Pro Leu Pro Leu Glu Arg Asp Cys
1 5 10 15

Sequence number: 8

Sequence length: 14

Type of sequence: amino acid

Topology: straight-chain

Class of sequence:

Sequence

Cys Arg Tyr Thr Asn Trp Ile Gln Lys Thr Ile Gln Ala Lys
1 5 10

Sequence number: 9

Sequence length: 26

Type of sequence: nucleic acid /14

Number of chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

CACAGAATTC CACCATGAAT CTACTT 26

Sequence number: 10

Sequence length: 27

Type of sequence: nucleic acid

Number of Chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

TAGCACCTGC CGATCTTGTC ATCATCA 27

Sequence number: 11

Sequence length: 28

Type of sequence: nucleic acid

Number of chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

GCAGACCTGC AGAACAAGTT GGTGCATG 28

Sequence number: 12

Sequence length: 18

Type of sequence: nucleic acid

Number of chains: one chain

Topology: straight-chain

Class of sequence: synthetic DNA

Sequence

AAAACCAGGG AGAATCAG 18

Brief Explanation of the Figures

Figure 1 is a photograph in lieu of a diagram of nitrocellulose film that shows results of attempting manifestation of gene SP59 in several types of human organs by Northern Blotting. PBL: peripheral blood lymphocyte

Figure 2 is a diagram showing results of studying enzyme activity of a mature protein that codes gene SP59 manifested in cell COS-1. The clear column shows when enterokinase was added, and the shaded column shows when enterokinase was not added. Moreover, pdKCR shows culture supernatant of cell COS-1 transfected only by the manifest vector used.

Photograph in Lieu of a Diagram

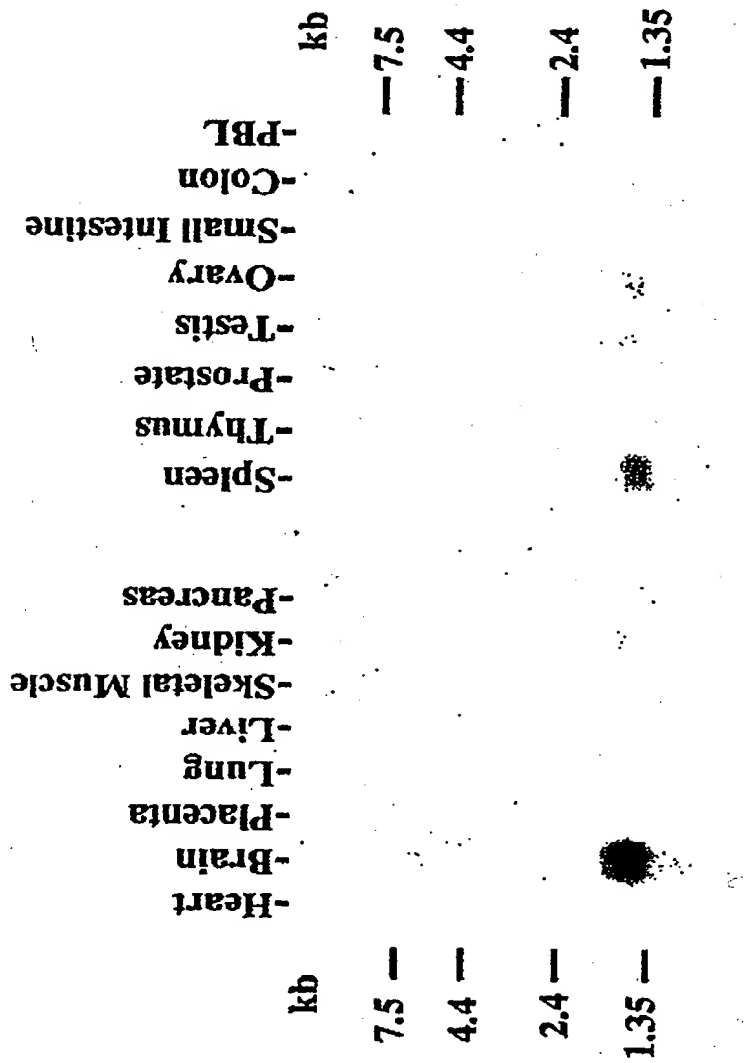

Figure 1

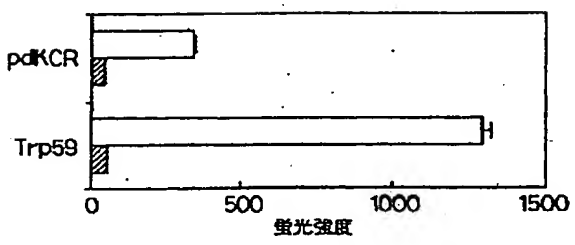

Figure 2
[below figure:] Fluorescent Intensity